Enterprise Risk Management Formula Book

1. Function definitions

Gamma, incomplete gamma, beta, incomplete beta, regularised incomplete beta, binomial

2. Series expansions

Exponential, natural logarithm, binomial, Taylor series

3. Calculus

Integration by parts, changing order of integration, differentiating an integral

4. Statistical distributions

Probability distribution terminology, Bayes' theorem, compound distributions

5. Statistical methods

Sample moments, parametric inference (with normal underlying distributions), maximum likelihood estimators, method-of-moments estimators, goodness of fit, linear regression, generalised linear models, correlations, analysis of variance, Bayesian priors and posteriors

6. Monte Carlo methods

Creation of normal random variables, Cholesky decomposition

7. Interest rates and bond pricing

Spot and forward rates, duration, modified duration, gross redemption yield (yield to maturity), credit spread, option-adjusted spread, annualisation conventions

8. Financial derivatives

Forward prices, Black-Scholes formulae

9. Risk measures

Value-at-Risk, tail Value-at-Risk, expected shortfall, expected worst outcome, tracking error, drawdown, marginal VaR, incremental VaR, estimating VaR

10. Portfolio optimisation

Mean-variance optimisation, capital asset pricing model

11. Extreme value theory

Maximum domain of attraction, Fisher-Tippett theorem, Pickands-Balkema-de Hann theorem, estimating tail distributions

12. Copulas

Definition, properties, Sklar's theorem, example copulas, tail dependence, simulating copulas

13. Miscellaneous

Combining solvency capital requirements using correlations, credit risk modelling, GARCH modelling, linear algebra and principal components, central limit theorem, Cornish-Fisher asymptotic expansion, Euler capital allocation principle, equiprobable outcomes for a multivariate normal distribution, RAROC and EVA

Appendix A: Probability distributions

Discrete: Binomial (and Bernoulli), Poisson

Continuous: Normal, uniform, chi-squared, exponential, *F*, generalised extreme value (GEV) (and Frechét, Gumbel and Weibull), generalised Pareto, lognormal, Student's *t*

Other: Distributional mixtures, location and scale adjusted distributions, multivariate distributions, distributional families

Tables: cumulative distribution function and quantile function for normal distribution

Note: In this note, $N(x)$ denotes the standard cumulative normal distribution function, $log\, x$ denotes logarithms to base e, $log^2 x = (log\, x)^2$ etc. but if $F(x)$ is a cumulative distribution function then $F^{-1}(x)$ is the corresponding inverse cumulative distribution function.

1. Function Definitions

[ERMFormulaBookFunctionDefinitions]

1.1 Gamma function, $\boldsymbol{\Gamma(q)}$

$$\Gamma(q) = \int_0^\infty t^{q-1}e^{-t}dt$$

Defined for $q \in \mathbb{R}$, q not a negative integer. Other properties: $\Gamma(q+1) = q\Gamma(q)$. If n is a positive integer then $n! = \Gamma(n+1)$. $\Gamma(1/2) = \sqrt{\pi}$.

1.2 Incomplete gamma function, $\boldsymbol{\Gamma_p(q)}$, beta function, $\boldsymbol{B(p,q)}$, incomplete beta function, $\boldsymbol{B_x(p,q)}$, regularised incomplete beta function, $\boldsymbol{I_x(p,q)}$

$$\Gamma_p(q) = \int_0^p t^{q-1}e^{-t}dt \qquad (p,q>0)$$

$$B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}dt \qquad (p,q>0)$$

$$B_x(p,q) = \int_0^x t^{p-1}(1-t)^{q-1}dt \qquad (p,q>0, 0 \le x \le 1)$$

$$I_x(p,q) = \frac{B_x(p,q)}{B(p,q)}$$

The beta function is related to the gamma function as follows:

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

The gamma, incomplete gamma, beta, incomplete beta and regularised incomplete beta can also be defined for negative (non-integral) values of p and q and for complex (non-real) values by analytic continuation.

1.3 The binomial coefficient, $\boldsymbol{\binom{n}{r}}$

For n and r integers ≥ 0 this is defined as:

$$\binom{n}{r} \equiv {}_nC_r \equiv \frac{n!}{(n-r)!\,r!} = \frac{\Gamma(n+1)}{\Gamma(n-r+1)\Gamma(r+1)} = \frac{n+1}{B(r+1, n-r+1)}$$

2. Series expansions (for real-valued functions)

[ERMFormulaBookSeriesExpansions]

2.1 Exponential function and natural logarithm (log) function

$$\exp(x) = e^x = 1 + x + \frac{x^2}{2!} + \cdots = \sum_{j=0}^{\infty} \frac{x^j}{j!}$$

$$\log(1+x) = \ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} \ldots = -\sum_{j=1}^{\infty} \frac{(-x)^j}{j}$$

2.2 Binomial expansion

$$(a+b)^n = a^n + \binom{n}{1} a^{n-1}b + \binom{n}{2} a^{n-2}b^2 + \cdots + b^n = \sum_{j=0}^{n} \binom{n}{j} a^{n-j}b^j$$

where $\binom{n}{r} = {}_nC_r = \frac{n!}{(n-r)!r!}$ is the binomial coefficient.

If we substitute into the binomial expansion $a = 1$, $b = x$ and $n = p$ we have (converges for any $p \in \mathbb{R}$ if $-1 < x < 1$):

$$(1+x)^p = 1 + px + \frac{p(p-1)}{2!}x^2 + \cdots$$

A corollary is that:

$$\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x \equiv \exp(x)$$

2.3 Taylor series expansion

For one variable: if series converges (where $f^{(j)}(x)$ is the j'th derivative of $f(x)$ and $f'(x) = f^{(1)}(x)$, $f''(x) = f^{(2)}(x)$ etc.):

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \cdots = \sum_{j=0}^{\infty} \frac{f^{(j)}(x)h^j}{j!}$$

For more than one variable: e.g. for two variables, if series converges (where $f'_x(x,y) = \frac{\partial f(x)}{\partial x}$, $f''_{xy}(x,y) = \frac{\partial^2 f(x)}{\partial x \partial y}$ etc.):

$$f(x+h, y+k) = f(x,y) + hf'_x(x,y) + kf'_y(x,y) + \frac{1}{2!}\left(h^2 f''_{xx}(x,y) + 2hkf''_{xy}(x,y) + h^2 f''_{yy}(x,y)\right) + \cdots$$

3. Calculus

[ERMFormulaBookCalculus]

3.1 Integration by parts

$$\int_a^b uv'dx = [uv]_a^b - \int_a^b vu'dx$$

3.2 Changing the order of integration in double integrals

Where the domain of integration is the set of values (x,y) for which $a \le y \le x \le b$:

$$\int_a^b \left(\int_a^x f(x,y)dy \right) dx = \int_a^b \left(\int_y^b f(x,y)dx \right) dy$$

or

$$\int_a^b dx \int_a^x dy\, f(x,y) = \int_a^b dy \int_y^b dx\, f(x,y)$$

3.3 Differentiating an integral

$$\frac{d}{dy} \int_{a(y)}^{b(y)} f(x,y)dx = b'(y)f(b(y),y) - a'(y)f(a(y),y) + \int_{a(y)}^{b(y)} \frac{\partial f(x,y)}{\partial y} dx$$

4. Statistical distributions

[ERMFormulaBookStatisticalDistributions]

4.1 Probability distribution terminology

Suppose a (continuous) real valued random variable, X, has a probability density function (or *pdf*) $p(x)$. Then the probability of X taking a value between x and $x + dx$ where dx is infinitesimal, $Pr(x \le X < x + dx)$, is $p(x)dx$.

The expected value of a function $f(x)$ (given this pdf) is defined (if the integral exists) as follows and is also sometimes written $\langle f(x) \rangle$:

$$E(f(X)) = \int_{-\infty}^{\infty} f(x)p(x)dx$$

For $p(x)$ to be a pdf it must exhibit certain basic regularity conditions including $\int_{-\infty}^{\infty} p(x)dx = 1$.

The mean, variance, standard deviation, cumulative distribution function (*cdf* or just *distribution function*), inverse cumulative distribution function (*inverse cdf* or just *inverse function* or *quantile function*), skewness (or *skew*), (excess) kurtosis, *mean excess function*, r'th *central* and *non-central moments* and *entropy* are defined as:

$$mean \equiv \bar{X}\ (= usually\ also\ \mu) \equiv E(X)$$
$$variance \equiv var(X) \equiv E\left((X - E(X))^2 \right)$$
$$standard\ deviation \equiv \sigma \equiv \sqrt{variance}$$
$$cumulative\ distribution\ function \equiv F(x)\ where\ F(x) = Pr(X \le x) = E(1_{\{X \le x\}}) = \int_{-\infty}^{x} p(y)dy$$
$$inverse\ cdf = quantile\ function = F^{-1}(q)\ where\ F^{-1}(F(x)) = x$$
$$skewness = skew(X) = E\left(\left(\frac{X - \mu}{\sigma} \right)^3 \right)$$
$$(excess)\ kurtosis = kurt(X) = E\left(\left(\frac{X - \mu}{\sigma} \right)^4 \right) - 3$$

$$mean\ excess\ function = e(u) = E(X - u|X > u)$$
$$characteristic\ function \equiv \varphi(t)\ where\ \varphi(t) = E\left(e^{itX}\right)$$
$$moment\ generating\ function \equiv M(t)\ where\ M(t) = E(e^{tX}) = \varphi(-it)$$
$$r'th\ central\ moment = \mu_r = E((X - \mu)^r)\quad usually\ r \in \mathbb{N}^+$$
$$r'th\ non-central\ moment = \mu_r' = E(X^r)\quad usually\ r \in \mathbb{N}^+$$
$$Entropy = E(-\log p(X))$$

The *cumulants* (sometimes called *semi-invariants*), κ_n, of a distribution, if they exist, are defined via the cumulant generating function, i.e. the power series expansion $\sum_{i=1}^{\infty} \kappa_n t^n/n!$ of $\log E(e^{tx})$. The mean, standard deviation, skewness and (excess) kurtosis of a distribution are $\mu = \kappa_1$, $\sigma^2 = \kappa_2$, $skew(X) = \kappa_3/\kappa_2^{3/2}$ and $kurt(X) = \kappa_4/\kappa_2^2$

The mode of a (continuous) distribution, i.e. $\arg\max_x p(x)$, is the value at which $p(x)$ is largest.

The median, *upper quartile* and *lower quartile* etc. (or more generally percentile) of a (continuous) distribution are $F^{-1}(0.5)$, $F^{-1}(0.75)$, $F^{-1}(0.25)$ etc. (or $F^{-1}(1-q)$) respectively.

Definitions of the above for discrete real-valued random variables are similar as long as the integrals involved are replaced with sums and the probability density function by the probability mass function $p(k) = Pr(X = k)$, i.e. the probability of X taking the value k.

Some of the above are not well defined or are infinite for some probability distributions.

If a discrete random variable can only take values which are non-negative integers, i.e. from the set $\{0,1,2,\dots\}$ then the probability generating function is defined as:

$$probability\ generating\ function = G(z) = E(z^X)$$

Characteristic functions and (if they exist) central moments and moment generating functions can nearly always be derived from non-central moments by applying the binomial expansion, e.g. $E(X+k) = E(X) + k$, $E((X+k)^2) = E(X^2) + 2kE(X) + k^2$ etc. (where k is a constant)

The *domain* (more fully, the *domain of definition* or *range*) of a (continuous) probability distribution is the set of values for which the probability density function is defined. The *support* of a (discrete) probability distribution is the set of values of x for which $Pr(X = x)$ is non-zero. The usual convention for a continuous function is to define the distribution only where the probability density function would be non-zero and for a discrete function (usually) to define the distribution only where the probability mass function is non-zero, in which case the domain/range and support coincide.

The *survival function* (or *reliability function*) is the probability that the variable takes a value greater than x (i.e. probability a unit survives beyond time x if x is measuring time) so is:

$$S(x) = P(X > x) = 1 - F(x)$$

The hazard function (also known as the *failure rate*) is the ratio of the pdf to the survival function, so is:

$$h(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{1 - F(x)}$$

The *cumulative hazard function* is the integral of the hazard function (i.e. the probability of failure at time x given survival to time x, if x is measuring time) so is:

$$H(x) = \int_{-\infty}^{x} h(t)dt = -\log\big(1 - F(x)\big)$$

Definitions, characteristics and common interpretations of a variety of (discrete and continuous) probability distributions are given in Appendix A.

The probability that Y occurs given that X occurs, $Pr(Y|X)$ is defined for $Pr(X) \neq 0$ as:

$$Pr(Y|X) = \frac{Pr(Y \cap X)}{Pr(X)} = \frac{Pr(Y,X)}{Pr(X)}$$

For discrete random variables, X, Y, the expected value of $f(Y)$ given that X occurs, $E(f(Y)|X)$ is defined as follows, where Q is the range of Y:

$$E(f(Y)|X = x) = \sum_{y \in Q} f(y)Pr(Y = y|X = x) = \sum_{y \in Q} f(y)\frac{Pr(Y = y, X = x)}{Pr(X = x)}$$

The following relationships apply:

$$E(Y) = E\big(E(Y|X)\big)$$
$$var(Y) = var\big(E(Y|X)\big) + E\big(var(Y|X)\big)$$

If $\mathbf{X} = (X_1, \dots, X_n)^T$ is a vector of (continuous) random variables then its (multivariate) pdf $f(x_1, \dots, x_n)$ and its cdf $F(x_1, \dots, x_n)$ satisfy:

$$Pr(x_1 \leq X_1 < x_1 + dx_1, \dots, x_n \leq X_n < x_n + dx_n) = \int \dots \int f(x_1, \dots, x_n)dx_1 \dots dx_n$$
$$Pr(X_1 \leq x_1, \dots, X_n \leq x_n) = F(x_1, \dots, x_n)$$

The covariance between X_i and X_j is $V_{ij} = cov(X_i, X_j) = E\Big(\big(X_i - E(X_i)\big)\big(X_j - E(X_j)\big)\Big)$ and the (Pearson) correlation coefficient is $\rho_{ij} = cov(X_i, X_j)/\sqrt{var(X_i)var(X_j)}$. The covariance matrix and the (Pearson) correlation matrix for multiple series are the matrices $\mathbf{V}$ and $\boldsymbol{\rho}$ which have as their elements V_{ij} and ρ_{ij} respectively.

4.2 Bayes theorem

Let $A_1, A_2, \dots, A_n$ be a collection of mutually exclusive and exhaustive events with probability of event A_j occurring being $P(A_j) \neq 0$ for $j = 1, \dots, n$. Then, for any event B such that $P(B) \neq 0$ the probability, $P(A_j|B)$, of A_j occurring conditional on B occurring (more simply the probability of A_j given B) satisfies:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{k=1}^{n} P(B|A_k)P(A_k)}$$

A *singly conditional* probability (i.e. order 1) is e.g. $P(A|B)$. A *doubly conditional* probability (i.e. order 2) is e.g. $P(A|B,C)$, probability of A occurring given both B and C take specific values. *Nil-conditioned* conditional probabilities (i.e. order 0) are the *marginal* probabilities, e.g. $P(A)$. A *Bayesian network* (more simply Bayesian net) is a directed acyclical graph where each node/vertex, say N_i is associated with a random variable, say X_i (often a two-valued, i.e. Boolean, random variable) and with a conditional probability table. For nodes without a parent the table contains just the marginal probabilities for the values that X_i might take. For nodes with parents it contains all conditional probabilities for the values that X_i might take given that its parents take specified values.

4.3 Compound distributions

If $X_1, X_2, \ldots$ are independent identically distributed random variables with moment generating function $M_X(t)$ and N is an independent non-negative integer-valued random variable then $S = X_1 + \cdots + X_N$ (with $S = 0$ when $N = 0$) has the following properties:

$$E(S) = E(N)E(X)$$
$$var(S) = E(N)var(X) + var(N)\big(E(X)\big)^2$$
$$M_S(t) = M_N(\log M_X(t))$$

For example, the compound Poisson distribution has: $E(S) = \lambda m_1$ and $var(S) = \lambda m_2$ where $\lambda = E(N)$ and $m_r = E(X^r)$

5. Statistical Methods

[ERMFormulaBookStatisticalMethods]

5.1 Sample moments

A random sample of n observations $(x_1, x_2, \ldots, x_n)$ has (equally weighted) sample moments as follows:

Sample mean	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Sample variance	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$
Sample skewness	$\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3$
Sample (excess) kurtosis	$\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$

'Population' moments (e.g. population variance, population skewness, population excess kurtosis) are calculated as if the distribution from which the data was being drawn was discrete and the probabilities of occurrence exactly matched the observed frequency of occurrence.

The least squares estimator for parameters of a distribution are the values of the parameters that minimise the square of the residuals, so the least squares estimator for the mean, $\hat{\mu}$, is the value that minimises $Y = \sum_{i=1}^{n}\big((x_i - \hat{\mu})\big)^2 \Rightarrow \frac{\partial Y}{\partial \hat{\mu}} = 0 \Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$

Non-equally weighted moments give different weights to different observations (the weights not dependent on the ordering of the observations), e.g. the sample non-equally weighted mean (using weights w_i) is:

$$\tilde{x} = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}$$

5.2 Parametric inference (with an underlying following the normal distribution)

One sample:

For a single (equally weighted) sample of size n, $(x_1, x_2, \dots, x_n)$, where $x_i \sim N(\mu, \sigma^2)$ then the following statistics are distributed according to the Student's *t* distribution and the chi-squared distribution:

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1} \quad and \quad \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

Two samples:

For two independent samples of sizes m and n, $(x_1, x_2, \dots, x_m)$ and $(y_1, y_2, \dots, y_n)$, where $x_i \sim N(\mu_x, \sigma_x^2)$ and $y_i \sim N(\mu_y, \sigma_y^2)$ then the following statistic is distributed according to the *F* distribution:

$$\frac{s_x^2/\sigma_x^2}{s_y^2/\sigma_y^2} \sim F_{m-1,n-1}$$

If $\sigma_x^2 = \sigma_y^2$ then:

$$\frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{s_p\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t_{m+n-2}$$

where $s_p^2 = \frac{1}{m+n-2}\left((m-1)s_x^2 + (n-1)s_y^2\right)$ is the pooled sample variance.

5.3 Maximum likelihood estimators

If $\hat{\theta}$ is the maximum likelihood estimator of a parameter θ based on a sample $\mathbf{X} = (X_1, \dots, X_n)$ then

$$\hat{\theta} = \arg\max_\theta L(\mathbf{X}|\theta)$$

where L is the likelihood for the sample, i.e. $L \equiv f(X_1|\theta) f(X_2|\theta) \cdots f(X_n|\theta)$ and hence $\log L\,(\mathbf{X}|\theta) = \sum_{i=1}^{n} \log f(X_i|\theta)$

$\hat{\theta}$ is asymptotically normally distributed with mean θ and variance equal to the Cramér-Rao lower bound

$$CRLB(\theta) = -\frac{1}{E\left(\frac{\partial^2}{\partial\theta^2}\log L(\theta, \mathbf{X})\right)}$$

Likelihood ratio test:

$$-2(l_p - l_{p+q}) = -2\log\left(\frac{\max_{H_0} L}{\max_{H_0 \cup H_1} L}\right) \sim^{approx} \chi_q^2$$

where $l_p = \max_{H_0} \log L$ is the maximum log-likelihood for the model under H_0 (with p free parameters) and $l_{p+q} = \max_{H_0 \cup H_1} \log L$ is the maximum log-likelihood for the model under $H_0 \cup H_1$ (with $p + q$ free parameters). Non-equally weighted estimators can be identified by weighting the $\log f(X_i|\theta)$ terms appropriately.

5.4 Method-of-moments estimators

Method of moments estimators are the parameter values (for the m parameters specifying a given distributional family) that result in replication of the first m moments of the observed data. For the normal distribution these involve $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$ and either $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = s^2$ (the sample variance, if a small sample size adjustment is included) or $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ (the 'population' variance, if the small sample size adjustment is ignored and we select the estimators to fit $E(X)$ and $E(X^2)$). In the *generalised method of moments* approach we select parameters that 'best' fit the selected moments (given some criterion for 'best'), rather than selecting parameters that perfectly fit the selected moments.

5.5 Goodness of fit

Goodness of fit describes how well a statistical model fits a set of observations. Examples include the following, where $x_{(i)}$ is the i'th order statistic, $\sup S$ is the supremum (i.e. largest value) of the set S, F is the cumulative distribution function of the distribution we are fitting and $F_n(x_{(i)})$ is the empirical distribution function:

(a) Kolmogorov-Smirnov test: $D_n = \sup_i \left|F_n(x_{(i)}) - F(x_{(i)})\right|$. Under the null hypothesis (that the sample comes from the hypothesized distribution), as $n \to \infty$ then $\sqrt{n}D_n$ tends to a limiting distribution (the Kolmogorov distribution).

(b) Cramér-von-Mises test: $T = n\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left(\frac{i}{2n} - F(x_{(i)})\right)^2$

(c) Anderson-Darling test: $A^2 = -n - S$ where $S = \sum_{i=1}^{n}\frac{2i-1}{n}\left(\log F(x_{(i)}) + \log\left(1 - F(x_{(n+1-i)})\right)\right)$

If data is bucketed into ranges then we may also use (Pearson's) chi-squared goodness of fit test using the following test statistic, where n is the sample size and O_i is the observed count, $E_i = n\left(F(Y_{i,u}) - F(Y_{i,l})\right)$ is the expected count and $Y_{i,l}$ and $Y_{i,u}$ are the lower and upper limits for the i'th bin. The test statistic follows approximately a chi-squared distribution with $k - c$ degrees of freedom, i.e. $\chi^2 \sim \chi_{k-c+1}^2$ where k is the number of non-empty cells and c is the number of estimated parameters plus 1:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

We may also test whether the skew or kurtosis or the two combined (the Jarque-Bera test) appear materially different from what would be implied by the relevant distributional family. If the null hypothesis is that the data comes from a normal distribution then, for large n, $skew \sim N(0, 6/n)$, $kurt \sim N(0, 24/n)$ and $JB = \frac{n}{6}\left(skew^2 + \frac{1}{4}kurt^2\right) \sim \chi_2^2$.

The Akaike Information Criterion (AIC) (and other similar ways of choosing between different types of model that trade-off goodness of fit with model complexity, such as the *Bayes Information Criterion*, BIC) involves selecting the model with the highest *information criterion* of the form $IC = \log L(\hat{\theta}) - f(n, q)$ where there are q unknown parameters and we are using a data series of length n for fitting purposes. For the AIC $f(n, q) = q$ and for the BIC $f(n, q) = q \log(n)/2$.

5.6 Linear regression

In the *univariate* case suppose $y_i = \alpha + \beta x_i + \varepsilon_i$ where $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \ldots, n$ then (equally weighted) estimates of α and β are:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad and \quad \hat{\beta} = \frac{s_{xy}}{s_{xx}}$$

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n-2}\left(s_{yy} - \frac{s_{xy}^2}{s_{xx}}\right)$$

where

$$s_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

$$s_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

$$s_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

Also

$$\frac{\hat{\beta} - \beta}{\sqrt{\frac{\hat{\sigma}^2}{s_{xx}}}} \sim t_{n-2}$$

The individual expected responses are $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$ and satisfy the following 'sum of squares' relationship:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

The variance of the predicted mean response is:

$$var(\hat{\alpha} + \hat{\beta}x) = \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{s_{xx}}\right)\sigma^2$$

The variance of a predicted individual response is the variance of the predicted mean response plus an additional σ^2.

For generalised least squares, if we have m different series each with n observations we are fitting $y_i = \sum_{i=1}^{m} \beta_j x_{ij} + \varepsilon_i$ then the vector of least squares estimators, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)^T$ is given by $\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ where $\mathbf{X}$ is a $n \times m$ matrix with elements x_{ij} and $\mathbf{Y}$ is an n dimensional vector with elements y_i.

5.7 Correlations

The observed (sample) correlation coefficient (i.e. *Pearson correlation coefficient*) between two series of equal lengths indexed in the same manner $(x_1, \dots, x_n)$ and $(y_1, \dots, y_n)$ is (where s_{xx}, s_{yy} and s_{xy} are as given in the section on linear regression):

$$r = s_{xy} / \sqrt{s_{xx}s_{yy}}$$

If the underlying correlation coefficient, ρ, is zero and the data comes from a bivariate normal distribution then:

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$$

For arbitrary ρ $(-1 < \rho < 1)$ the Fisher z transform is $z(r)$ where:

$$z(r) = tanh^{-1}(r) = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$$

If the data comes from a bivariate normal distribution then $z(r)$ is distributed approximately as follows:

$$z(r) \sim N\left(z(\rho), \frac{1}{n-3}\right) \quad (approximately)$$

Two non-parametric measures of correlation are:

- Spearman's rank correlation coefficient, where q_t and r_t are the ranks within x and y of x_t and y_t respectively:

$$\rho_{Spearman} = \frac{\sum_{t=1}^{n}(q_t - \bar{q})(r_t - \bar{r})}{\sqrt{\sum_{t=1}^{n}(q_t - \bar{q})^2 . \sum_{t=1}^{n}(r_t - \bar{r})^2}} \quad where\ \bar{q} = \frac{1}{n}\sum_{t=1}^{n} q_t\ etc.$$

- Kendall's tau, where computation is taken over all $i = 1,2, \dots, n$ and $j = 1,2, \dots, n$ with $i \neq j$ and (for the moment ignoring ties) a concordant pair is a case where $(x_i > x_j\ and\ y_i > y_j)\ or\ (x_i < x_j\ and\ y_i < y_j)$ and a discordant pair is a case where $(x_i > x_j\ and\ y_i < y_j)\ or\ (x_i < x_j\ and\ y_i > y_j)$:

$$\tau = \frac{(number\ of\ concordant\ pairs) - (number\ of\ discordant\ pairs)}{\frac{1}{2}n(n-1)}$$

There are various possible ways of handling ties in these two non-parametric measures of correlation (ties should not in practice arise if the random variables really are continuous).

5.8 Analysis of variance

Given a single factor normal model

$$y_{ij} \sim N(\mu + q_i, \sigma^2), \quad i = 1,2,\dots,k \quad j = 1,2,\dots,n_i$$

where $n = \sum_{i=1}^{k} n_i$ with $\sum_{i=1}^{k} n_i q_i = 0$.

Variance estimate:

$$\hat{\sigma}^2 = \frac{SS_R}{n-k}$$

Under the null hypothesis given above

$$\frac{SS_B/(k-1)}{SS_R/(n-k)} \sim F_{k-1,n-k}$$

where:

$$SS_T = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} {y_{ij}}^2 - \frac{\bar{y}^2}{n}$$

$$SS_B = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{k}\frac{\bar{y}_i^2}{n_i} - \frac{\bar{y}^2}{n}$$

$$SS_R = SS_T - SS_B$$

$$\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} \quad and \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}$$

5.9 Bayesian priors and posteriors

Posterior and prior distributions are related as follows:

$$Posterior \propto Prior \times Likelihood$$

i.e.

$$p(\theta|x) \propto p(\theta) \times p(x|\theta)$$

For example, if x is a random sample of size n from a $N(\mu, \sigma^2)$ where σ^2 is known and the prior distribution for μ is $N(\mu_0, \sigma_0^2)$ then the posterior distribution for μ is:

$$\mu|x \sim N(\mu_*, \sigma_*^2)$$

where μ_* is 'credibility weighted' as follows:

$$\mu_* = \frac{\left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right)}{\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)}$$

and

$$\sigma_*^2 = \frac{1}{\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)}$$

6. Monte Carlo methods

[ERMFormulaBookMonteCarloMethods]

6.1 Creation of normal random variables

Box-Muller: if x_1 and x_2 are independent standard uniform random variables, i.e. come from $U(0,1)$ and $r = \sqrt{-2\log x_1}$ then $x_1 = r\cos(2\pi x_2)$ and $u_2 = r\sin(2\pi x_2)$ are independent standard normal random variables.

Polar method: if u_1 and u_2 are independent random variables from $U(-1,1)$, $r^2 = u_1^2 + u_2^2$ and $s = \sqrt{-2\log(r^2)/r^2}$ then $x_1 = su_1$ and $x_2 = su_2$ are independent standard normal random variables.

6.2 Cholesky decomposition

If A has real entries, is symmetric and is positive definite then it can be decomposed as $A = LL^T$ where L is a lower triangular matrix with strictly positive diagonal entries and L^T is its transpose. The entries of L are:

$$L_{j,j} = \sqrt{A_{j,j} - \sum_{k=1}^{j-1} L_{j,k}^2} \qquad L_{i,j} = \frac{1}{L_{j,j}}\left(A_{i,j} - \sum_{k=1}^{j-1} L_{i,k}L_{j,k}\right) \quad for\ i > j$$

7. Interest rates and bond pricing

[ERMFormulaBookIRBondPricing]

7.1 Spot and forward rates

Suppose $P(t)$ is the price at time 0 of a zero-coupon bond that pays 1 at time t, $s(t)$ is the spot rate for the period $(0,t)$, i.e. 0 to t, and $f(t)$ is the instantaneous forward rate at time 0 for time t (where $s(t)$ and $f(t)$ are both continuously compounded) Then:

$$P(t) = e^{-ts(t)} = \exp\big(-ts(t)\big) = \exp\left(-\int_0^t f(u)du\right)$$

$$s(t) = \frac{1}{t}\log P(t)$$

$$f(t) = -\frac{d}{dt}\log P(t)$$

7.2 Duration, modified duration, (gross) redemption yield (yield to maturity), credit spread, option-adjusted spread, annualisation conventions

If a bond gives the holder entitlements to cash flows C_t at time t (and is assumed not to be subject to default risk) and has a 'dirty price', V, then its (gross) redemption yield (yield to maturity) is the (sensible) rate of interest that equates V with its present value, i.e.:

$$PV = \sum_t \frac{C_t}{(1+i)^t}$$

Its duration is then $dur = (1/V)\sum_t tC_t/(1+i)^t$ and its modified duration is $mod\,dur = -(1/V)\,dV/di = dur/(1+i)$. $V \times mod\,dur$ is almost exactly the same as its PV01, also called DV01.

Its credit spread is the difference between its (gross) redemption yield and the corresponding yield on a reference security, often a corresponding government security providing the same cash flows in the event of non-default. The option-adjusted spread is the corresponding spread taking into account optionality in the bond in question and/or in the reference bond.

Interest rates may be expressed as annual rates, i, semi-annual rates, $i^{(2)}$, quarterly rates, $i^{(4)}$, monthly rates, $i^{(12)}$ or even continuously compounded rates, δ, where:

$$1+i = \left(1+\frac{i^{(2)}}{2}\right)^2 = \left(1+\frac{i^{(4)}}{4}\right)^4 = \left(1+\frac{i^{(12)}}{12}\right)^{12} = \lim_{n\to\infty}\left(1+\frac{\delta}{n}\right)^n = \exp\delta$$

The quotation convention of a bond (e.g. ACT/ACT) defines the amount of accrued interest payable when a bond is bought or sold in between coupon dates.

8. Financial derivatives

[ERMFormulaBookFinancialDerivatives]

8.1 Forward prices

The no arbitrage (fair) forward price which parties should agree to exchange a security at time T if it is priced S_0 now and is entitled to fixed income of present value Q in the meantime is:

$$F = (S_0 - I)e^{rT}$$

where r is the interest rate (continuously compounded).

If instead it pays dividends at a rate q (continuously compounded) then the forward prices is:

$$F = S_0 e^{(r-q)T}$$

8.2 Black-Scholes formulae

Geometric Brownian motion for a security (stock) price S_t involves

$$\frac{dS_t}{S_t} = d(\log S_t) = \mu dt + \sigma dz$$

The partial differential equation satisfied by values of payoffs involving such security prices is:

$$\frac{\partial V}{\partial t} + (r - q)S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = rV$$

where r is the interest rate, q is the dividend yield (both continuously compounded) and σ is the security price volatility.

Garman-Kohlhagen formulae for values at time t of European-style put and call options with strike price K maturing at time T:

Call option $$C_t = S_t e^{-q(T-t)} N(d_1) - K e^{-r(T-t)} N(d_2)$$

Put option $$P_t = K e^{-r(T-t)} N(-d_2) - S_t e^{-q(T-t)} N(-d_1)$$

where

$$d_1 = \frac{\log(S_t/K) + (r - q + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$$
$$d_2 = \frac{\log(S_t/K) + (r - q - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} = d_1 - \sigma\sqrt{T-t}$$

We then have $C_t + K e^{-r(T-t)} = P_t + S_t e^{-q(T-t)}$, i.e. *put-call parity*.

Technically the Black-Scholes formulae are special cases of the Garman-Kohlhagen formulae for stocks that pay no dividend, i.e. have $q = 0$, although in practice the two names are normally treated as interchangeable.

The Black-Scholes option pricing formulae can also be derived as the limit of binomial trees (lattices) with movements $S \to S(1+u)$ or $S \to S(1-d)$ with an up-step probability $p_{up} = \left(e^{r\Delta t} - d\right)/(u-d)$ and a down-step probability $p_{down} = 1 - p_{up}$ where:

$$u \approx e^{\sigma\sqrt{\Delta t} + q\Delta t} \quad and \quad d \approx e^{-\sigma\sqrt{\Delta t} + q\Delta t}$$

9. Risk measures

[ERMFormulaBookRiskMeasures]

9.1 Value-at-Risk (VaR)

If X is a (continuous) random variable (e.g. an outcome) with pdf $p(x)$ then the Value-at-Risk at confidence level α (e.g. 95%, 99%, 99.5%) is defined as:

$$VaR_\alpha(X) = k \quad where \int_{-\infty}^{-k} p(x)dx = 1 - \alpha$$

If X has cdf $F(x)$ with an inverse cdf, i.e. quantile function $F^{-1}(q)$ then $VaR_\alpha(X) = -F^{-1}(\alpha)$. Sometimes signs are inverted and α and $1-\alpha$ are swapped around when defining $VaR(\alpha)$.

The *relative VaR* of X relative Y, e.g. of an active equity portfolio versus a benchmark portfolio is usually taken to mean the VaR of the random variable $Z = X - Y$. However, for relative returns there are several alternative ways in which we can define the equivalent of $X - Y$, see definition of tracking error.

9.2 Tail Value-at-Risk (TVaR)

The Tail Value-at-Risk (also called the *conditional Value-at-Risk*, CVaR) is generally defined as the value of the loss conditional on it being worse than the VaR at confidence level α, so is defined as:

$$TVaR_\alpha(X) = E(-X|X \leq -VaR_\alpha) = -\frac{1}{1-\alpha}\int_{-\infty}^{-VaR_\alpha} xp(x)dx$$

A coherent risk measure is one that satisfies *subadditivity*, *monotonicity*, *homogeneity* and *translational invariance*. If losses follow a continuous probability distribution then TVAR is a coherent risk measure.

Occasionally TVaR (less commonly CVaR) is ascribed the same meaning as expected shortfall, in which case the $1/(1-\alpha)$ factor is ignored, or is defined relative to some specific limit $-k$ that in effect defines the α to be used in the above formula.

9.3 Expected shortfall (ES)

The expected shortfall, ES, and expected policyholder deficit, EPD are usually defined as follows:

Expected policyholder deficit:

$$EPD(W) = -E\big((X-W)I(X<W)\big) = -\int_{-\infty}^{W}(x-W)p(x)dx$$

where $I(X<W) = \begin{cases} 1, & X < W \\ 0, & X \geq W \end{cases}$ and W is often but not always the policyholder wealth

Expected shortfall:

$$ES = -\int_{-\infty}^{0} xp(x)dx$$

Or more generally the expected shortfall below some trigger level Q is

$$ES(Q) = -\int_{-\infty}^{Q} xp(x)dx$$

Sometimes expected shortfall is ascribed the same meaning as is given above for TVaR.

9.4 Expected worst outcome (EWO)

The expected value of the worst outcome in N (non-overlapping) observations is:

$$EWO(N) = \int_{-\infty}^{\infty} x_{(1)} dF^N$$

where the integral is N-dimensional, $x_{(1)} = \min(x_1, \dots, x_n)$ and the joint distribution F^N involves N independent marginal distributions each with pdf $p(x)$. This type of risk measure can also be extended to, say, n'th worst outcome, $n \le N$ with $EWO(N)$ as defined above being the special case where $n = 1$.

9.5 Tracking error (TE)

If X is a random variable (e.g. a portfolio return) with (assumed forward looking) pdf $p(x)$ then its *ex-ante tracking error* (if it exists) is σ where

$$\sigma^2 = var(X)$$

Nearly always X is here the relative return of a portfolio $\mathbf{p} = (p_1, \dots, p_n)^T$ of exposures versus a benchmark portfolio $\mathbf{b} = (b_1, \dots, b_n)^T$ and the tracking error is then normally expressed as a percentage of the total portfolio value. The tracking error of $\mathbf{p}$ versus $\mathbf{b}$ is then $\sigma(\mathbf{p}, \mathbf{b})^2$ (more precisely, $\sigma_t(\mathbf{p}, \mathbf{b})^2$ for a time period indexed by t) where if the future returns on the i'th instrument during this are $r_{i,t}$ and relative returns are calculated arithmetically (i.e. using an arithmetic difference):

$$\sigma_t(\mathbf{p}, \mathbf{b})^2 = var\left(\sum_{i=1}^{n} p_i r_{i,t} - \sum_{i=1}^{n} b_i r_{i,t}\right) = var\left(\sum_{i=1}^{n} (p_i - b_i) r_{i,t}\right) = var\left(\sum_{i=1}^{n} a_i r_{i,t}\right) = \sigma_t(\mathbf{a})^2$$

where $\mathbf{a}$ is the vector of active positions.

If the $r_i(t)$ have covariance matrix $\mathbf{V}$ with elements V_{ij} then $\sigma(\mathbf{a})^2 = \mathbf{a}^T \mathbf{V} \mathbf{a}$.

However, returns compound rather than add through time so for non-infinitesimal time period lengths there are alternative and potentially preferable ways of defining relative returns, including (if we are trying to calculate the return r_1 relative to r_2, each expressed as fractions) using geometric relative returns, i.e. $r_{geometric\ relative} = (1 + r_1)/(1 + r_2) - 1$, or logarithmic relative returns, i.e. $r_{logarithmic\ relative} = \log(1 + r_1) - \log(1 + r_2)$ rather than arithmetic relative returns $r_{arithmetic\ relative} = r_1 - r_2$.

If a factor structure is assumed for the $r_{i,t}$ then this normally involves assuming that:

$$r_{i,t} = \alpha_i + \sum_{k} \beta_{i,j} x_{j,t} + \varepsilon_{i,t}$$

where $\beta_{i,j}$ is the exposure (beta) of the i'th instrument to the j'th factor and $\varepsilon_{i,t}$ are residual (idiosyncratic) components.

A portfolio described by a vector of (active) weights $\mathbf{a}$ then has an expected return of $\mathbf{a}.\boldsymbol{\mu}$ and an (expected) future tracking error as follows, where β is the matrix formed by $\beta_{i,j}$ and $\widetilde{\mathbf{V}}$ is the covariance matrix between the factors

$$\sigma^2 = \mathbf{a}^T\left(\boldsymbol{\beta}^T\widetilde{\mathbf{V}}\boldsymbol{\beta}\right)\mathbf{a} + idiosyncratic\ components$$
$$= (\boldsymbol{\beta}\mathbf{a})^T\widetilde{\mathbf{V}}(\boldsymbol{\beta}\mathbf{a}) + idiosyncratic\ components$$

If T is the length of time (time horizon) to which the ex-ante tracking error relates and returns in individual time periods of length t are assumed to be independent of each other then (assuming e.g. we measure returns logarithmically and that the portfolio and benchmark remain unchanged through time) we can apply the *square-root of time* adjustment to derive ex ante tracking errors applicable to different time periods, i.e.:

$$\sigma_T = \sigma_t\sqrt{(T/t)}$$

A portfolio's *ex-post tracking error* is derived from past observed values of its returns and might then be either a sample standard deviation or (perhaps less accurately, but slightly lower) the corresponding 'population' standard deviation.

9.6 Drawdown

If a portfolio has exhibited past returns $r_1, \dots, r_n$ over the previous n time periods (which could be days, weeks, months, years etc., where $t = 1$ is earlier than $t = 2$ etc.) then the portfolio's *drawdown* at time t is usually defined to be r_t (if negative). Its *maximum drawdown* is usually defined as $\min(r_1, \dots, r_t)$ (if negative). Its *cumulative maximum drawdown* (i.e. *peak-to-trough*) at time t is usually defined by creating an index I_t such that $I_t = I_{t-1}(1 + r_t)$ and then determining at time t the maximum of $(I_{u-k} - I_u)/I_{u-k}$ for all $u \le t$ and $k < u$.

9.7 Marginal VaR

The overall outcome of a portfolio of exposures containing a (constant) amount a_i of exposure to the i'th risk where each risk involves an random outcome X_i (technically X_i is the value ascribed to the random outcome) is $X = \sum_i a_i X_i$. Strictly speaking combining exposures in this manner requires that the way in which we ascribe a financial value to an outcome satisfies the axioms of uniqueness, additivity and scalability, i.e. that $V(X)$, the value we ascribe to an outcome should be unique and should satisfy $V\big(k(A + B)\big) = k\big(V(A) + V(B)\big)$.

The VaR of such a portfolio with confidence level α is $VaR_\alpha(X) = VaR_\alpha(\sum_i a_i X_i)$.

The marginal VaR with confidence level α of the i'th exposure in such a portfolio is:

$$MVaR_\alpha^{(i)} = \frac{\partial}{\partial a_i}\left(VaR_\alpha\left(\sum_i a_i X_i\right)\right)$$

The contribution to overall VaR of the i'th exposure is then $a_i MVaR_\alpha^{(i)}$.

If the outcomes are Gaussian (i.e. multivariate normal, say m exposures with $\mathbf{X} = (X_1, \dots, X_m)^T \sim N(\boldsymbol{\mu}, \mathbf{V})$ then $X = \sum_i a_i X_i \sim N(\mu, \sigma^2)$ where:

$$\mu = \sum_{i=1}^{m} a_i \mu_i = \mathbf{a}.\boldsymbol{\mu}$$

$$\sigma^2 = \sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j V_{ij} = \mathbf{a}^T\mathbf{V}\mathbf{a}$$

Here $\mathbf{a} = \left(a_{1,} \dots, a_m\right)^T$ and $\boldsymbol{\mu} = \left(\mu_{1,} \dots, \mu_m\right)^T$. Given the properties of the normal distribution

$$VaR_\alpha(X) = \mu + \sigma N^{-1}(1-\alpha)$$

As tracking error is a special case of VaR (with assumed normal underlying distribution and $\mu = 0$, $N^{-1}(1-\alpha) = 1$) we can likewise define the marginal tracking error and contribution to tracking error from an individual exposure.

9.8 Incremental VaR

The incremental VaR with confidence level α of the i'th exposure in such a portfolio is:

$$IVaR_\alpha^{(i)} = VaR_\alpha\left(\sum_i a_i X_i\right) - VaR_\alpha\left(\sum_{j,j\neq i} a_j X_j\right)$$

9.9 Estimating VaR

If the observations are normally distributed then VaR_α may be estimated approximately in a parametric manner using $\widehat{VaR}_\alpha = \hat{\mu} + \hat{\sigma}N^{-1}(1-\alpha)$. Alternatively it can be estimated (approximately) in a non-parametric manner (if the data does not exhibit temporal dependencies) by taking the observations $Y_1, \dots, Y_n$, say, reordering them so that $Y_{(1)} \leq \dots \leq Y_{(n)}$, say, identifying the r'th order statistic, where r is an integer between 1 and n, and estimating the VaR using the k'th order statistic where $(k-1)/n < (1-\alpha) \leq k/n$. Using a binomial distribution, the variance of the r'th order statistic is approximately as follows (where $f_{(r)}$ is the pdf at $y_{(r)}$ and p is the probability of outcome) meaning that estimating the standard error of this non-parametric statistic requires us to estimate $f_{(r)}$:

$$var\left(y_{(r)}\right) \approx \frac{p(1-p)}{nf_{(r)}^2}$$

10. Portfolio optimisation

[ERMFormulaBookPortfolioOptimisation]

10.1 Mean-variance portfolio optimisation

If there are n asset categories then a (one period) mean-variance efficient portfolio, $\mathbf{x} = (x_1, \dots, x_n)^T$ where x_i is the amount (or weight) invested in the i'th asset, given a benchmark, $\mathbf{b} = (b_1, \dots, b_n)^T$, assumed future (one period) mean returns on each asset, $\mathbf{r} = (r_1, \dots, r_n)^T$, and an assumed covariance matrix between the (one period) returns on different assets, $\mathbf{V}$, is a portfolio that maximises the utility function, $U(\mathbf{x})$ for some risk aversion parameter, λ, subject to relevant constraints on the x_i, where:

$$U(\mathbf{x}) = \mathbf{r}.\mathbf{x} - \lambda\left((\mathbf{x}-\mathbf{b})\right)^T\mathbf{V}(\mathbf{x}-\mathbf{b})$$

The constraints that are applied are usually linear, i.e. of the form $\mathbf{Ax} \leq \mathbf{q}$ which if $\mathbf{q}$ is an m dimensional vector is understood as meaning that there are m constraints each of the form $\sum_{i=1}^{n} A_{ij}x_i \leq q_j$. In such a formulation, equality constraints, including that the amounts invested add up to the total portfolio value (or the weights add to unity), can be written as two inequality constraints, e.g. $\sum_{i=1}^{n} x_i \leq 1$ and $\sum_{i=1}^{n}(-1)x_i \leq -1$ combined.

The implied alphas with a mean-variance risk-return model given a portfolio $\mathbf{x}$ and benchmark $\mathbf{b}$ are the mean returns that need to be assumed for the different asset categories for $\mathbf{x}$ to be mean-variance optimal for some value of λ. They can only meaningfully be determined for assets whose weights in the portfolio are not constrained (other than by the constraint that weights add to unity). They are then $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^T$ where $\boldsymbol{\alpha} = c_1 + c_2\lambda\mathbf{V}(\mathbf{x} - \mathbf{b})$ where c_1 and c_2 are arbitrary scalar constants.

10.2 Capital Asset Pricing Model (CAPM)

The security market line is:

$$E_i - r = \beta_i(E_M - r)$$

where $\beta_i = cov(R_i, R_M)/var(R_M)$

The capital market line (for efficient portfolios) is:

$$E_P - r = (E_M - r)\frac{\sigma_M}{\sigma_M}$$

11. Extreme value theory

[ERMFormulaBookExtremeValueTheory]

See also here.

11.1 Maximum domain of attraction (MDA)

Suppose that i.i.d. random variables X_i have cdf $F(x)$. Suppose also that there exist sequences $\{c_i\}$ and $\{d_i\}$ and a cdf $H(x)$ such that:

$$\lim_{n\to\infty} Pr\left(\frac{M_n - d_n}{c_n} \leq x\right) = H(x)$$

where M_n is the random variable corresponding to the block maximum for blocks of such variables of length n, i.e. each (independent) realisation of the series $\{X_i\}$ is used to create a realisation of M_n given by $M_n = \max(X_1, \dots, X_n)$.

Then F is said to be in the *maximum domain of attraction* (MDA) of H, written $F \in MDA(H)$

11.2 Fisher-Tippett theorem

If $F \in MDA(H)$ where H is a non-degenerate cdf then H must be a Generalised Extreme Value (GEV) distribution.

If $F \in MDA(H_{\xi,\mu,\sigma})$ where $H = GEV(\xi,\mu,\sigma)$ then by replacing c_i by $\tilde{c}_i = \sigma c_i$ and d_i by $\tilde{d}_i = d_i + \mu c_i$ we see that $F \in MDA(H_\xi)$ where $H = GEV(\xi)$.

11.3 The Pickands-Balkema-de Haan (PBH) theorem

Let x_F be the maximum limiting value of the random variable X. Then the PBH theorem states that we can find a function $\beta(u)$ such that

$$\lim_{u \to x_F} \left(\sup_{0 \le y < x_F - u} \left| F_u(y) - G_{\xi,\beta(u)}(y) \right| \right) = 0$$

if and only if that $F \in MDA(H_\xi)$

11.4 Estimating tail distributions

Suppose that the underlying loss distribution is in the maximum domain of attraction of the Frechét distribution and it has a tail of the form: $\bar{F}(x) = L(x)x^{-\alpha}$ for some slowly varying $L(x)$, where $\bar{F}(x) = 1 - F(X)$. Then the Hill estimator for the (upper) tail index, given n ordered observations, $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$, assuming that the (upper) tail contains k entries is:

$$\hat{\alpha}_{k,n} = \left(\frac{1}{k} \sum_{j=1}^{k} \left(\log X_{(n-j+1)} - \log X_{(n-k)} \right) \right)^{-1}$$

12. Copulas

[ERMFormulaBookCopulas]

12.1 Definition

A copula is a multivariate cumulative distribution function for an n dimensional random vector $U = (U_1, \dots, U_n)^T$ in the unit hypercube ($[0,1]^n$) that has uniform marginals, U_i, each distributed according to $U(0,1)$ but not in general independent of each other. Let $u = (u_1, \dots, u_n)^T$ also be restricted to the unit hypercube $[0,1]^n$. Then a copula is defined as a function of the form:

$$C(u) = C(u_1, \dots, u_n) = Pr(U_1 \le u_1, \dots, U_n \le u_n)$$

Equivalently $C(u_1, \dots, u_n)$ is the joint cumulative distribution function for the random vector $U \in [0,1]^n$.

The *copula density* (for a continuous copula) is the pdf for which the cdf is the copula.

12.2 Properties

In the bivariate case ($n = 2$) for a general function $C(u_1, u_2)$ to be a copula it must satisfy the following properties:

1. $C(u, 1) = u = C(1, u)$ for all $0 \le u \le 1$
2. $C(u_1, u_2)$ must be increasing in both u_1 and u_2

3. $C(b_1, b_2) - C(a_1, b_2) - C(b_1, a_2) + C(a_1, a_2) \geq 0$ for all $0 \leq a_1 < b_1 \leq 1$ and $0 \leq a_2 < b_2 \leq 1$
4. $C(u_1, u_2) \leq \min(u_1, u_2)$
5. $C(u_1, u_2) \geq \max(u_1 + u_2 - 1, 0)$

12.3 Sklar's theorem

If F is a joint (cumulative) distribution with marginal cdf's $F_1, F_2, \ldots, F_n$ then there exists a copula C which maps the unit hypercube $[0,1]^n$ onto the interval $[0,1]$ such that for all $x_1, \ldots, x_n$ we have:

$$F(x_1, \ldots, x_n) = C\big(F_1(x_1), \ldots, F_n(x_n)\big)$$

Moreover, if the F_i are continuous functions then the copula is unique and

$$C(u_1, \ldots, u_n) = F\left(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)\right)$$

Conversely, suppose $C(u_1, \ldots, u_n)$ is a copula and $F_1(x_1), \ldots, F_n(x_n)$ are univariate cdf's. Then the function $F(x_1, \ldots, x_n) = C\big(F_1(x_1), \ldots, F_n(x_n)\big)$ is a joint distribution function with marginal cdf's $F_1, F_2, \ldots, F_n$.

12.4 Example copulas

The Archimedean family involves copulas of the following form, where $\varphi: [0,1] \to [0, \infty)$, $\varphi(0) = \infty$, $\varphi(1) = 0$, φ is continuous and strictly decreasing and $(-1)^k \, d^k \varphi^{-1}(t)/dt^k \geq 0 \quad \forall k = 0,1, \ldots$

$$C(u_1, \ldots, u_n) = \varphi^{-1}\big(\varphi(u_1) + \cdots + \varphi(u_n)\big)$$

Special cases include the Clayton copula which has $\varphi(t) = t^{-\theta} - 1$ (for some suitable value of θ) and the independence or product copula which has $\varphi(t) = -\log t$.

12.5 Tail dependence

If X_1 and X_2 are continuous random variables with copula $C(u_1, u_2)$ then their coefficient of (joint lower) tail dependence (if it exists) is:

$$\lambda \equiv \lim_{u \to 0} \frac{C(u, u)}{u}$$

For continuous random variables X and Y each with lower limit of $-\infty$ the coefficient of (lower) tail dependence is also:

$$\lambda = \lim_{z \to -\infty} Pr(Y < z | X < z) = \lim_{z \to -\infty} \frac{Pr(Y < z, X < z)}{Pr(X < z)}$$

12.6 Simulating copulas

Correlated Gaussian (i.e. multivariate normal) random variables (i.e. random variables with a Gaussian copula and Gaussian marginals) can be generated using Cholesky decomposition.

For random variables that have a Gaussian copula but non-normal marginal (with cdfs $F_1, \dots, F_n$) we can generate a vector $(x_1, \dots, x_n)^T$ of correlated Gaussian random variables as above and then transform as per $y_i = F_i^{-1}(x_i)$.

In general, for non-Gaussian copulas we may need to generate a vector of unit uniform random variables $(u_1, \dots, u_n)^T$ and then transform them using $u_1^* = u_1$, $u_2^* = C^{-1}(u_2|u_1)$ etc.

13. Miscellaneous

[ERMFormulaBookMiscellaneous]

13.1 Combining solvency capital requirements using correlations

A correlation based combination of individual solvency capital requirements involves a formula along the lines of:

$$SCR_{Tot} = \sqrt{\sum_{i,j} c_{ij} \times SCR_i \times SCR_j}$$

13.2 Credit risk modelling

A single factor credit portfolio model generally assumes $z_k = \alpha_k x + \varepsilon_k \sqrt{1 - \alpha_k^2}$ where f is the (standardised) factor return/movement, α_k is the exposure of the k'th obligor to that factor and ε_k is the idiosyncratic noise term for that obligor.

If $\alpha_k^2 = \rho$ is the same for all instruments (e.g. all are assumed to have same correlation with market plus only an idiosyncratic term) then $z_k = \sqrt{\rho} x + \varepsilon_k \sqrt{1 - \rho}$. In such circumstances and if all obligors have the same probability of default p say then probability $p_{k,n}$ that k out of n default is:

$$P_k = \binom{n}{k} \int_{-\infty}^{\infty} \left(s(u)\right)^k \left(1 - s(u)\right)^{n-k} dN(u) \quad where\ s(u) = N\left(\frac{1}{\sqrt{1-\rho}}\left(N^{-1}(p) - u\sqrt{\rho}\right)\right)$$

If $F_n(\theta)$ is the cumulative probability that the percentage loss on the portfolio does not exceed θ then in the well diversified limit (Vasicek's loss distribution):

$$F_\infty(\theta) = \lim_{n \to \infty} F_n(\theta) = N\left(\frac{1}{\sqrt{\rho}}\left(N^{-1}(\theta)\sqrt{1-\rho} - N^{-1}(p)\right)\right)$$

13.3 GARCH models

Risk models may cater for heteroscedasticity by including GARCH features, e.g. the model might involve formulae along the lines of (in practice μ will slowly evolve as additional data is received):

$$x_{t+1} = \mu + \sigma_t \varepsilon_t$$
$$\varepsilon_t \sim N(0,1)$$

where for, say, a GARCH(1,1) model

$$\sigma_t^2 = \alpha + \beta(x_t - \mu)^2 + \gamma\sigma_{t-1}^2$$

RiskMetrics typically uses the following approach for estimating σ_t^2 (often using $\lambda_i = \lambda^i$ for some suitably chosen decay factor λ) which can be viewed as an example of a GARCH approach and/or using weighted moments.

$$\sigma_t^2 = \frac{\sum_{i=0}^{t-1} \lambda_i (x_{t-i} - \mu)^2}{\sum_{i=1}^{t} \lambda_i}$$

13.4 Linear algebra and principal components

Suppose we have $i = 1, \dots, m$ data series (e.g. returns) each with $j = 1, \dots, n$ observations, $X_{i,j}$, that are coincident in time across the different data series. Suppose the $m \times m$ covariance matrix of the (empirical) covariances between the different series is $\mathbf{V}$. The *eigenvalues* and *eigenvectors* of V are the values of λ (scalar) and associated $\mathbf{x}$ (vector) for which $\mathbf{Vx} = \lambda\mathbf{x}$. An $m \times m$ matrix has m (not necessarily distinct) eigenvalues and associated eigenvectors. Eigenvectors associated with distinct eigenvalues are *orthogonal*, i.e. $\mathbf{x}_i^T \mathbf{x}_k = 0$ for $i \neq k$. *Orthonormal* eigenvectors have $|\mathbf{x}_i| = \mathbf{x}_i^T \mathbf{x}_i = 1$ and $\mathbf{x}_i^T \mathbf{x}_k = 0$ for $i \neq k$. For any distinct eigenvalue the associated orthonormal eigenvector is unique up to a change of sign. If $q > 1$ eigenvalues all take the same value then it is possible to find q orthogonal eigenvectors corresponding to all of these eigenvalues. For empirical covariance matrices, $\mathbf{V}$ is symmetric non-negative definite (and positive definite if no two data series are perfectly correlated) and all of its m eigenvalues, λ_i, are greater than or equal to zero. One way of telling if a matrix is positive definite is to test whether it is possible to apply a Cholesky decomposition to it.

The eigenvalues and associated eigenvectors of an empirical covariance matrix may be sorted so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$. The *first principal component* is the mixture of the underlying (de-meaned) series, i.e. the $r_j = \sum_{q=1}^{m} b_q \left(X_{q,j} - \bar{X}_i\right)$, that corresponds to the orthonormal eigenvector, $\mathbf{b}$, corresponding to the largest eigenvalue of V. This choice of $\mathbf{b}$ maximises $\mathbf{b}^T\mathbf{Vb}$ subject to $|\mathbf{b}| = 1$, Other (lesser) principal components correspond to orthonormal eigenvectors corresponding to smaller eigenvalues.

13.5 Central limit theorem

Suppose we have a series of independent random variables $X_1, \dots, X_n, \dots$ each with finite (bounded) expected value μ_i and finite (bounded) standard deviation σ_i. Suppose S_n and Z_n are defined as:

$$S_n^2 = \sum_{i=1}^{n} \sigma_i^2 \qquad Z_n = \frac{1}{S_n} \sum_{i=1}^{n} (X_i - \mu_i)$$

Then subject to certain regularity conditions the distribution of Z_n tends asymptotically to $N(0,1)$ (it is exactly $N(0,1)$ if each of the X_i is normally distributed).

13.6 Cornish-Fisher asymptotic expansion

The (4th moment) Cornish-Fisher asymptotic expansion approximates a standardised QQ-plot via the following function:

$$y_{CF4}(x) = x + \frac{\gamma_1(x^2-1)}{6} + \frac{3\gamma_2(x^3-3x) - 2\gamma_1^2(2x^3-5x)}{72}$$

where γ_1 and γ_2 are the skew and excess kurtosis of the data.

13.7 Euler capital allocation principle

A function $f(\mathbf{u})$ where $\mathbf{u} = (u_1, \dots, u_n)^T$ is said to be homogenous of order q if:

$$\sum_{i=1}^{n} u_i \left(\left. \frac{\partial f}{\partial u_i} \right|_{k\mathbf{u}} \right) = q k^{q-1} f(\mathbf{u})$$

Suppose we have n business lines, the outcome (loss) to each business line given its current size is L_i (a random variable) so the total loss is $L = \sum_{i=1}^{n} u_i L_1$ where for the current business portfolio the business line allocation is $\mathbf{u} = (u_1, \dots u_n)^T = (1, \dots, 1)^T = \mathbf{1}$. Suppose the risk measure used to determine economic capital is $r(x)$ and that it is homogeneous of order 1, i.e. $r(kL) = kr(L)$. Then the Euler capital allocation principle (and, in effect, the Marginal VaR or Internal beta approach to setting RAROC rates) allocates total economic capital, EC (technically a function of the business portfolio allocation, $\mathbf{u}$) into capital for each business line, EC_i, as follows:

$$EC = \sum_{i=1}^{n} EC_i \quad where\ EC_i = \left. \left(u_i \frac{\partial EC(\mathbf{u})}{\partial u_i} \right) \right|_{\mathbf{u}=\mathbf{1}}$$

13.8 Equiprobable outcomes for a multivariate normal distribution

If $\mathbf{X} \sim N(\boldsymbol{\mu}, \mathbf{V})$ where $\boldsymbol{\mu} = \mathbf{0}$ then equiprobable scenarios (i.e. contours where $p(X)$ is constant) are ellipsoids defined by $\mathbf{x}^T \mathbf{V}^{-1} \mathbf{x} = k^2$ for some constant value of k. The probability that X lies within this ellipsoid is given by a chi-squared with n degrees of freedom:

$$Pr(X\ inside\ ellipsoid) = \chi_n^2(k^2)$$

13.9 RAROC, EIC, SHV and SVA

Risk adjusted return on capital (RAROC) is usually defined as follows, where E = *Adjusted earnings* = *Earnings – Interest cost – expected loss – funding cost – other costs* and K = *capital*:

$$RAROC = \frac{E}{K}$$

Economic income created (EIC) is usually defined as where c = *per unit cost of equity* (i.e. *hurdle rate*):

$$EIC = E - c \times K$$

Shareholder value (SHV) and shareholder value added (SVA) (also known as economic value added, EVA) translate current period return contribution to overall economic value. Given suitable assumptions about future growth prospects for a business, g, these are:

$$SHV = K \times \left(\frac{RAROC - g}{c - g}\right) \qquad SVA = K \times \left(\frac{RAROC - g}{c - g} - 1\right)$$

Appendix A: Probability Distributions

[ERMFormulaBookAppendix]

Definitions, characteristics, tabulations and common interpretations of a variety of (discrete and continuous) probability distributions are given below. Please note that some probability distributions have multiple names or have special cases called different names.

For further details of these and other distributions recognised by the Nematrian website see here.

A.1: Discrete (univariate) distributions

[ERMFormulaBookAppendixDiscrete]

Binomial (and Bernoulli), Poisson

Distribution name	Binomial distribution
Common notation	$X \sim B(n,p)$
Parameters	n = number of (independent) trials, positive integer p = probability of success in each trial, $0 \le p \le 1$
Support	$x \in \{0,1,\dots,n\} = number\ of\ successes$
Probability mass function	$f(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{(n-x)!\,x!} p^x (1-p)^{n-x}$
Cumulative distribution function	$F(x) = \sum_{j=0}^{x} \binom{n}{j} p^j (1-p)^{n-j} = I_{1-p}(n-x, x+1)$
Mean	np
Variance	$np(1-p)$
Skewness	$\frac{1-2p}{\sqrt{np(1-p)}}$
(Excess) kurtosis	$\frac{1-6p(1-p)}{np(1-p)}$
Characteristic function	$\left(1-p+pe^{it}\right)^n$
Other comments	Corresponds to the number of successes in a sequence of n independent experiments each of which has a probability p of being successful. The *Bernoulli* distribution is $B(1,p)$ and corresponds to the likelihood of success of a single experiment. Its probability mass function and cumulative distribution function are:

	$$f(x) = F(x) = \begin{cases} 1-p, & x=0 \\ p, & x=1 \end{cases}$$ The Bernoulli distribution with $p = 1/2$, i.e. $B(1, 1/2)$, has the minimum possible excess kurtosis, i.e. -2. The mode of $B(n,p)$ is $int\big((n+1)p\big)$ if $(n+1)p$ is 0 or not an integer and is n if $(n+1)p = n+1$. If $(n+1)p \in \{1,2,\dots,n\}$ then the distribution is bi-modal, with modes $(n+1)p$ and $(n+1)p-1$.

Distribution name	Poisson distribution
Common notation	$X \sim Pois(\lambda)$
Parameters	λ = event rate ($\lambda > 0$)
Support	$x \in \{0,1,2,\dots\}$
Probability mass function	$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$
Cumulative distribution function	$$F(x) = e^{-\lambda} \sum_{j=0}^{x} \frac{\lambda^j}{j!}$$ (can also be expressed using the incomplete gamma function)
Mean	λ
Variance	λ
Skewness	$\lambda^{-1/2}$
(Excess) kurtosis	λ^{-1}
Characteristic function	$e^{\lambda(e^{it}-1)}$
Other comments	Expresses the probability of a given number of events occurring in a fixed interval of time if the events occur with a known average rate and independently of the time since the last event. The median is approximately $int(\lambda + 1/3 - 0.02/\lambda)$. The mode is $int(\lambda)$ if λ is not integral. Otherwise the distribution is bi-modal with modes λ and $\lambda - 1$.

A.2: Continuous (univariate) distributions

[ERMFormulaBookAppendixContinuous]

- normal, uniform, chi-squared: see here
- exponential, *F*, generalised extreme value (GEV) (and Frechét, Gumbel and Weibull): see here
- generalised Pareto, lognormal, Student's *t*: see here

(a) Normal, uniform, chi-squared

[Nematrian website page: ERMFormulaBookAppendixContinuous1, © Nematrian 2017]

Distribution name	Normal distribution
Common notation	$X \sim N(\mu, \sigma^2)$

Parameters	σ = scale parameter ($\sigma > 0$) μ = location parameter
Domain	$-\infty < x < +\infty$
Probability density function	$f(x) \equiv \phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$
Cumulative distribution function	$F(x) \equiv N(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right) dt$
Mean	μ
Variance	σ^2
Skewness	0
(Excess) kurtosis	0
Characteristic function	$e^{it\mu-\frac{1}{2}\sigma^2 t^2}$
Other comments	The normal distribution is also called the *Gaussian* distribution. The *unit normal* (or *standard normal*) distribution is $N(0,1)$. The inverse unit normal distribution function (i.e. its quantile function) is commonly written $N^{-1}(x)$ (also in some texts $\Phi(x)$ and the unit normal density function is commonly written $\phi(x)$. $N^{-1}(x)$ is also called the *probit* function. The *error function distribution* is $N\left(0,\frac{1}{2h}\right)$, where h is now an inverse scale parameter $h > 0$. The median and mode of a normal distribution are μ. The truncated first moments of $N(\mu,\sigma^2)$ are: $\int_{L}^{U} xf(x)dx = \mu\left(N\left(\frac{U-\mu}{\sigma}\right) - N\left(\frac{L-\mu}{\sigma}\right)\right) - \sigma\left(\phi\left(\frac{U-\mu}{\sigma}\right) - \phi\left(\frac{L-\mu}{\sigma}\right)\right)$ where $\phi(x)$ and $N(x)$ are the pdf and cdf of the unit normal distribution respectively. The mean excess function of a standard normal distribution is thus $e(u) = \frac{\phi(u) - uN(-u)}{N(-u)} = \frac{\frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}u^2\right) - uN(-u)}{N(-u)}$ The central moments of the normal distribution are: $E\left((X-\mu)^k\right) = \begin{cases} 0 & if\ k\ is\ odd \\ \sigma^k \times 1 \times 3 \times \ldots \times (k-1) & if\ k\ is\ even \end{cases}$

Distribution name	Uniform distribution
Common notation	$X \sim U(a,b)$
Parameters	a, b = boundary parameters ($a < b$)

Domain	$a \leq x \leq b$
Probability density function	$f(x) = \frac{1}{b-a}$
Cumulative distribution function	$F(x) = \frac{x-a}{b-a}$
Mean	$(a+b)/2$
Variance	$(b-a)^2/12$
Skewness	0
(Excess) kurtosis	$-6/5$
Characteristic function	$\frac{e^{ibt} - e^{iat}}{it(b-a)}$
Other comments	Its non-central moments $(r = 1,2,3,\ldots)$ are $E(X^r) = \frac{1}{(b-1)}\frac{1}{r+1}(b^{r+1} - a^{r+1})$. Its median is $(a+b)/2$.

Distribution name	Chi-squared distribution
Common notation	$X \sim \chi_\nu^2$
Parameters	ν = degrees of freedom (positive integer)
Domain	$0 \leq x < +\infty$
Probability density function	$f(x) = \frac{x^{\nu/2-1}\exp\left(-\frac{x}{2}\right)}{2^{\nu/2}\Gamma(\nu/2)}$
Cumulative distribution function	$F(x) = \frac{\Gamma_{x/2}(\nu/2)}{\Gamma(\nu/2)}$
Mean	ν
Variance	2ν
Skewness	$2\sqrt{\frac{2}{\nu}}$
(Excess) kurtosis	$\frac{12}{\nu}$
Characteristic function	$(1-2it)^{-\nu/2}$
Other comments	Its median is approximately $\nu\left(1 - \frac{2}{9\nu}\right)^3$. Its mode is $\max(\nu - 2, 0)$. Is also known as the central chi-squared distribution (when there is a need to contrast it with the noncentral chi-squared distribution). In the special case of $\nu = 2$ the cumulative distribution function simplifies to $F(x) = 1 - e^{-x/2}$. The chi-squared distribution with ν degrees of freedom is the distribution of a sum of the squares of ν independent standard normal random variables. A consequence is that the sum of independent chi-squared variables is also chi-squared distributed. It is widely used in hypothesis testing, goodness of fit analysis or in constructing confidence intervals. It is a special case of the gamma distribution. As $\nu \to \infty$, $(\chi_\nu^2 - \nu)/\sqrt{2\nu} \to N(0,1)$ and $qF(q,\nu) \to \chi_q^2$

(b) exponential, F, generalised extreme value (GEV) (and Frechét, Gumbel and Weibull)

[ERMFormulaBookAppendixContinuous2]

Distribution name	Exponential distribution
Common notation	$X \sim Exp(\lambda)$
Parameters	λ = inverse scale (i.e. rate) parameter ($\lambda > 0$)
Domain	$0 \leq x < +\infty$
Probability density function	$f(x) = \lambda \exp(-\lambda x)$
Cumulative distribution function	$F(x) = 1 - \exp(-\lambda x)$
Mean	$\frac{1}{\lambda}$
Variance	$\frac{1}{\lambda^2}$
Skewness	2
(Excess) kurtosis	6
Characteristic function	$(1 - i\,t/\lambda)^{-1}$
Other comments	Also called the *negative exponential* distribution. The mode of an exponential distribution is 0. The exponential distribution describes the time between events if these events follow a Poisson process. It is not the same as the exponential family of distributions. The quantile function, i.e. the inverse cumulative distribution function, is $F^{-1}(p;\lambda) = -\frac{\log(1-p)}{\lambda}$. The non-central moments ($r = 1,2,3,\dots$) are $E(X^r) = \frac{\Gamma(1+r)}{\lambda^r}$. Its median is $\frac{\log 2}{\lambda}$.

Distribution name	*F* distribution
Common notation	$X \sim F(\nu_1, \nu_2)$
Parameters	ν_1 = degrees of freedom (first) (positive integer) ν_2 = degrees of freedom (second) (positive integer)
Domain	$0 \leq x < +\infty$
Probability density function	$f(x) = \frac{1}{xB(\nu_1/2, \nu_2/2)} \sqrt{\frac{(\nu_1 x)^{\nu_1} {\nu_2}^{\nu_2}}{(\nu_1 x + \nu_2)^{\nu_1+\nu_2}}}$
Cumulative distribution function	$F(x) = \frac{B_{(\nu_1 x)/(\nu_1 x + \nu_2)}(\nu_1, \nu_2)}{B(\nu_1/2, \nu_2/2)} = I_{(\nu_1 x)/(\nu_1 x + \nu_2)}(\nu_1/2, \nu_2/2)$
Mean	$\frac{\nu_1}{\nu_2 - 2}$ for $\nu_2 > 2$
Variance	$\frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$ for $\nu_2 > 4$
Skewness	$\frac{(2\nu_1 + \nu_2 - 2)\sqrt{8(\nu_2 - 4)}}{(\nu_2 - 6)\sqrt{\nu_1(\nu_1 + \nu_2 - 2)}}$ for $\nu_2 > 6$
(Excess) kurtosis	$12\frac{\nu_1(5\nu_2 - 22)(\nu_1 + \nu_2 - 2) + (\nu_2 - 4)(\nu_2 - 2)^2}{\nu_1(\nu_2 - 6)(\nu_2 - 8)(\nu_1 + \nu_2 - 2)}$ for $\nu_2 > 8$

Characteristic function	$$\frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_2}{2}\right)}U\left(\frac{\nu_1}{2},1-\frac{\nu_2}{2},-\frac{\nu_2}{\nu_1}it\right)$$ Where $U(a,b,z)$ is the confluent hypergeometric function of the second kind
Other comments	The *F* distribution is a special case of the Pearson type 6 distribution. It is also known as Snedecor's *F* or the Fisher-Snedecor distribution. It commonly arises in statistical tests linked to analysis of variance. If $X_1 \sim \chi^2(\nu_1)$ and $X_2 \sim \chi^2(\nu_2)$ are independent random variables then $$\frac{X_1/\nu_1}{X_2/\nu_2} \sim F(\nu_1,\nu_2)$$ The *F*-distribution is a particular example of the beta prime distribution. The mode is $\frac{(\nu_1-2)}{\nu_1}\frac{\nu_2}{\nu_2+2}$ for $\nu_1 > 2$. There is no simple closed form for the median.

Distribution name	Generalised extreme value (GEV) distribution (for maxima)
Common notation	$X \sim GEV(\xi,\mu,\sigma)$
Parameters	ξ = shape parameter μ = location parameter σ = scale parameter
Domain	$$1+\left(\frac{x-\mu}{\sigma}\right)\xi > 0 \quad \xi \neq 0$$ $$-\infty < x < \infty \quad \xi = 0$$
Probability density function	$$f(x) = \frac{1}{\sigma}Q(x)^{\xi+1}e^{-Q(x)}$$ where $$Q(x) = \begin{cases} \left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi} & \xi \neq 0 \\ \exp\left(-\frac{x-\mu}{\sigma}\right) & \xi = 0 \end{cases}$$
Cumulative distribution function	$$F(x) = e^{-Q(x)}$$
Mean	$$\begin{cases} \mu+\sigma\frac{\Gamma(1-\xi)-1}{\xi} & if\ \xi \neq 0, \xi < 1 \\ \mu+\sigma\gamma & if\ \xi = 0 \\ \infty & \xi \geq 1 \end{cases}$$ where γ is Euler's constant, i.e. $\lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k}-\log n\right)$
Variance	$$\begin{cases} \sigma^2\frac{g_2-g_1^2}{\xi^2} & if\ \xi \neq 0, \xi < 1/2 \\ \frac{\sigma^2\pi^2}{6} & if\ \xi = 0 \\ \infty & \xi \geq 1/2 \end{cases}$$

	Where $g_k = \Gamma(1-k\xi)$
Skewness	$\begin{cases} \dfrac{g_3 - 3g_1g_2 + 2g_1^3}{(g_2 - g_1^2)^{3/2}} & if\ \xi \neq 0 \\ \dfrac{12\sqrt{6}\varsigma(3)}{\pi^3} & if\ \xi = 0 \end{cases}$ where $\varsigma(x)$ is the Riemann zeta function, i.e. $\sum_{k=1}^{\infty}\frac{1}{k^x}$.
(Excess) kurtosis	$\begin{cases} \dfrac{g_4 - 4g_1g_3 + 6g_2g_1^2 - 3g_1^4}{(g_2 - g_1^2)^2} & if\ \xi \neq 0 \\ \dfrac{12}{5} & if\ \xi = 0 \end{cases}$
Other comments	ξ defines the tail behaviour of the distribution. The sub-families defined by $\xi = 0$ (Type I), $\xi > 0$ (Type II) and $\xi < 0$ (Type III) correspond to the Gumbel, Frechét and Weibull families respectively. An important special case when analysing threshold exceedances involves $\mu = 0$ (and normally $\xi > 0$) and this special case may be referred to as $GEV(\xi, \sigma)$.

(c) generalised Pareto, lognormal, Student's *t*

[ERMFormulaBookAppendixContinuous3]

Distribution name	Generalised Pareto distribution (GPD)
Common notation	$X \sim GPD(\xi, \mu, \sigma)$
Parameters	ξ = shape parameter μ = location parameter σ = scale parameter ($\sigma > 0$)
Domain	$\mu \leq x < +\infty \quad \xi \geq 0$ $\mu \leq x \leq \mu - \dfrac{\sigma}{\xi} \quad \xi < 0$
Probability density function	$f(x) = \begin{cases} \dfrac{1}{\sigma}(1+\xi z)^{-1-1/\xi} & \xi \neq 0 \\ \dfrac{1}{\sigma}\exp(-z) & \xi = 0 \end{cases}$ where $z = \dfrac{x-\mu}{\sigma}$
Cumulative distribution function	$F(x) = \begin{cases} 1-(1+\xi z)^{-1/\xi} & \xi \neq 0 \\ 1-\exp(-z) & \xi = 0 \end{cases}$
Mean	$\mu + \dfrac{\sigma}{1-\xi} \quad \xi < 1$
Variance	$\dfrac{\sigma^2}{(1-2\xi)(1-\xi)^2} \quad \xi < \dfrac{1}{2}$
Skewness	$\dfrac{2(1+\xi)\sqrt{1-2\xi}}{1-3\xi} \quad \xi < \dfrac{1}{3}$
(Excess) kurtosis	$\dfrac{6(1+\xi-6\xi^2-2\xi^3)}{(1-3\xi)(1-4\xi)} \quad \xi < \dfrac{1}{4}$
Other comments	GPD is used in the peaks over thresholds variant of extreme value theory

Distribution name	Lognormal distribution
Common notation	$X \sim logN(\mu, \sigma^2)$
Parameters	σ = scale parameter ($\sigma > 0$) μ = location parameter
Domain	$0 < x < +\infty$
Probability density function	$f(x) = \dfrac{\exp\left(-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right)}{x\sigma\sqrt{2\pi}}$
Cumulative distribution function	$F(x) = N\left(\frac{\log x - \mu}{\sigma}\right) = \frac{1}{2} + \frac{1}{2} erf\left(\frac{\log x - \mu}{\sqrt{2\sigma^2}}\right)$
Mean	$e^{\mu + \sigma^2/2}$
Variance	$\left(e^{(\sigma^2)} - 1\right)e^{2\mu + \sigma^2}$
Skewness	$\left(e^{(\sigma^2)} + 2\right)\sqrt{e^{(\sigma^2)} - 1}$
(Excess) kurtosis	$e^{(4\sigma^2)} + 2e^{(3\sigma^2)} + 3e^{(2\sigma^2)} - 6$
Characteristic function	No simple expression that is not divergent
Other comments	The median of a lognormal distribution is e^{μ} and its mode is $e^{\mu - \sigma^2}$. The truncated moments of $logN(\mu, \sigma^2)$ are: $\int_L^U x^k f(x)dx = e^{k\mu + k^2\sigma^2/2}\left(N\left(\frac{\log U - \mu}{\sigma} - k\sigma\right) - N\left(\frac{\log L - \mu}{\sigma} - k\sigma\right)\right)$

Distribution name	(Standard) Student's *t* distribution
Common notation	$X \sim t_{\nu}$
Parameters	ν = degrees of freedom ($\nu > 0$, usually ν is an integer although in some situations a non-integral ν can arise)
Domain	$-\infty < x < +\infty$
Probability density function	$f(x) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} = \frac{1}{\sqrt{\nu}B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$
Cumulative distribution function	$F(x) = \begin{cases} \frac{1}{2} I_z\left(\frac{\nu}{2}, \frac{1}{2}\right) & x < 0 \\ 1 - \frac{1}{2} I_z\left(\frac{\nu}{2}, \frac{1}{2}\right) & x \geq 0 \end{cases}$ where $z = \nu/(\nu + x^2)$
Mean	0
Variance	$\frac{\nu}{\nu - 2} \quad for\ \ \nu > 2$
Skewness	$0 \quad for\ \nu > 3$
(Excess) kurtosis	$\frac{3(\nu - 2)}{\nu - 4} \quad for\ \ \nu > 4$
Characteristic function	$\frac{K_{\nu/2}\left(\sqrt{\nu}\lvert t\rvert\right)\left(\sqrt{\nu}\lvert t\rvert\right)^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2-1}}$ where $K_{\nu/2}(x)$ is a Bessel function

Other comments	The *Student's t* distribution (more simply the *t* distribution) arises when estimating the mean of a normally distributed population when sample sizes are small and the population standard deviation is unknown. It is a special case of the generalised hyperbolic distribution. Its non-central moments if r is even and $0 < r < \nu$ are: $$E(X^r) = \frac{\Gamma\left(\frac{r+1}{2}\right)\Gamma\left(\frac{\nu-r}{2}\right)\nu^{r/2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}$$ If r is even and $0 < \nu \leq r$ then $E(X^r) = \infty$, if r is odd and $0 < r < \nu$ then $E(X^r) = 0$ and if r is odd and $0 < \nu \leq r$ then $E(X^r)$ is undefined.

A.3: Distributional mixtures

[ERMFormulaBookAppendixMixtures]

A random variable is distributed according to a *normal mixture distribution* if it is of the form $X = m(W) + \sqrt{W}\sigma Z$ where W and Z are independent random variables, W is a non-negative random variable, $Z \sim N(0,1)$ and $m(W)$ is some function of W. For example, the *t* distribution has $m(W) = \mu$ and $1/W$ being chi-squared with ν degrees of freedom and the standard non-central t distribution has $m(W) = d\sqrt{W}$ where d is the non-centrality parameter and $1/W$ is chi-squared with ν degrees of freedom.

A *distributional mixture of normal distributions* is to be interpreted more generally as any distribution in which the overall random variable is selected with probability p_i from a (typically finite) number of normal random distributions, the i'th one of which is $N(\mu_i, \sigma_i^2)$ for arbitrary constant μ_i and σ_i. Any univariate distribution can be approximated arbitrarily accurately with a large enough number of underlying normal random distributions. It is contrasted with a *linear combination mixture* of normal distributions in which the overall random variable is derived by adding together a linear combination of underlying normal random variables, i.e. $X = a_1 X_1 + \cdots + a_n X_n$.

A.4: Location and scale adjusted distributions

[ERMFormulaBookAppendixLocationScale]

The location and scale of any probability distribution can be adjusted by using the (linear) transform $Y = g + hX$ where g and h are constants. This leaves the skew and (excess kurtosis) unaltered but alters the mean and variance as follows: $E(Y) = g + hE(X)$ and $var(Y) = h^2 var(X)$.

In some cases, the typical distributional specification already includes such components. For example, the normal distribution $N(\mu, \sigma^2)$ is the location and scale adjusted version of the unit normal distribution $N(0,1)$.

In other cases, the standard distributional specification does not include such adjustments. For example, the (standard) Student's *t* distribution depends one just one parameter, its degrees of freedom. In these cases, the distributional definition noted above may need to be expanded to include location and/or scale adjusted variants when fitting data to such distributions.

A.5: Multivariate probability distributions

[ERMFormulaBookAppendixMultivariate]

Multivariate normal (i.e. Gaussian) distribution

The multivariate probability distribution $N(\boldsymbol{\mu}, \mathbf{V})$ where $\boldsymbol{\mu}$ is a vector of n elements and $\mathbf{V}$ is an $n \times n$ non-negative definite matrix has the following joint density function (where $|\mathbf{V}| = \det \mathbf{V}$ is the determinant of V)

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{|\mathbf{V}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{V}(\mathbf{x}-\boldsymbol{\mu})\right)$$

The means of the individual marginal distributions are μ_i where $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$ and the covariance between the i'th and the j'th marginal distributions are V_{ij} where the V_{ij} are the elements of $\mathbf{V}$. Its moment generating function is $M(t) = \exp\left(\mathbf{t}^T\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^T\mathbf{V}\mathbf{t}\right)$ and its characteristic function is $\varphi(t) = \exp\left(i\mathbf{t}^T\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^T\mathbf{V}\mathbf{t}\right)$. The multivariate normal distribution has as its copula the Gaussian copula.

A bivariate random variable $X = (X_1, X_2)^T$ follows a *standard bivariate normal* distribution if it has $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{V} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. More generally, a multivariate normal distribution is a *standard multivariate normal* distribution if $\boldsymbol{\mu} = \mathbf{0}$ and a covariance matrix which is also a correlation matrix, i.e. where the variance of each individual marginal distribution is 1.

For numerical values of the cumulative distribution function of the standard bivariate normal distribution see here.

A.6: Distributional families

[ERMFormulaBookProbabilityDistributions]

Exponential family

A (continuous) random variable $\mathbf{X}$ (multivariate or univariate) follows a distribution from the exponential family, with vector parameter $\boldsymbol{\theta}$, if its probability (density) function can be written in the form:

$$f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta}) = h(\mathbf{x}) \exp\big(\boldsymbol{\eta}(\boldsymbol{\theta})\mathbf{T}(\mathbf{x}) - A(\boldsymbol{\theta})\big)$$

or equivalently as:

$$f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta}) = h(\mathbf{x})g(\boldsymbol{\theta}) \exp\big(\boldsymbol{\eta}(\boldsymbol{\theta}).\mathbf{T}(\mathbf{x})\big)$$

The $g(\boldsymbol{\theta})$ are automatically determined once the other functions have been chosen because $\int f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})d\mathbf{x} = 1$. $\boldsymbol{\eta}$ is called the natural parameter.

A.7: Standard (i.e. unit) normal distribution:

(a) Cumulative distribution function

[ERMFormulaBookAppendixCumulativeNormal]

The table below tabulates the cumulative distribution function of the unit normal distribution, i.e.

$$N(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}t^2\right) dt$$

- for values of x from 0.00 to 1.00 see here
- for values of x from 1.00 to 2.00 see here
- for values of x from 2.00 to 3.00 see here
- for values of x from 3.00 to 4.00 see here

(1)

[ERMFormulaBookAppendixCumulativeNormal1]

The table below tabulates the cumulative distribution function of the unit normal distribution, i.e.

$$N(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}t^2\right) dt$$

for values of x from 0.00 to 1.00:

x	$N(x)$	x	$N(x)$	x	$N(x)$	x	$N(x)$	x	$N(x)$
0.00	0.50000	**0.20**	0.57926	**0.40**	0.65542	**0.60**	0.72575	**0.80**	0.78814
0.01	0.50399	**0.21**	0.58317	**0.41**	0.65910	**0.61**	0.72907	**0.81**	0.79103
0.02	0.50798	**0.22**	0.58706	**0.42**	0.66276	**0.62**	0.73237	**0.82**	0.79389
0.03	0.51197	**0.23**	0.59095	**0.43**	0.66640	**0.63**	0.73565	**0.83**	0.79673
0.04	0.51595	**0.24**	0.59483	**0.44**	0.67003	**0.64**	0.73891	**0.84**	0.79955
0.05	0.51994	**0.25**	0.59871	**0.45**	0.67364	**0.65**	0.74215	**0.85**	0.80234
0.06	0.52392	**0.26**	0.60257	**0.46**	0.67724	**0.66**	0.74537	**0.86**	0.80511
0.07	0.52790	**0.27**	0.60642	**0.47**	0.68082	**0.67**	0.74857	**0.87**	0.80785
0.08	0.53188	**0.28**	0.61026	**0.48**	0.68439	**0.68**	0.75175	**0.88**	0.81057
0.09	0.53586	**0.29**	0.61409	**0.49**	0.68793	**0.69**	0.75490	**0.89**	0.81327
0.10	0.53983	**0.30**	0.61791	**0.50**	0.69146	**0.70**	0.75804	**0.90**	0.81594
0.11	0.54380	**0.31**	0.62172	**0.51**	0.69497	**0.71**	0.76115	**0.91**	0.81859
0.12	0.54776	**0.32**	0.62552	**0.52**	0.69847	**0.72**	0.76424	**0.92**	0.82121
0.13	0.55172	**0.33**	0.62930	**0.53**	0.70194	**0.73**	0.76730	**0.93**	0.82381
0.14	0.55567	**0.34**	0.63307	**0.54**	0.70540	**0.74**	0.77035	**0.94**	0.82639
0.15	0.55962	**0.35**	0.63683	**0.55**	0.70884	**0.75**	0.77337	**0.95**	0.82894
0.16	0.56356	**0.36**	0.64058	**0.56**	0.71226	**0.76**	0.77637	**0.96**	0.83147
0.17	0.56749	**0.37**	0.64431	**0.57**	0.71566	**0.77**	0.77935	**0.97**	0.83398
0.18	0.57142	**0.38**	0.64803	**0.58**	0.71904	**0.78**	0.78230	**0.98**	0.83646
0.19	0.57535	**0.39**	0.65173	**0.59**	0.72240	**0.79**	0.78524	**0.99**	0.83891
0.20	0.57926	**0.40**	0.65542	**0.60**	0.72575	**0.80**	0.78814	**1.00**	0.84134

(2)

[ERMFormulaBookAppendixCumulativeNormal2]

The table below tabulates the cumulative distribution function of the unit normal distribution, i.e.

$$N(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}t^2\right) dt$$

for values of x from 1.00 to 2.00:

x	$N(x)$	x	$N(x)$	x	$N(x)$	x	$N(x)$	x	$N(x)$
1.00	0.84134	**1.20**	0.88493	**1.40**	0.91924	**1.60**	0.94520	**1.80**	0.96407
1.01	0.84375	**1.21**	0.88686	**1.41**	0.92073	**1.61**	0.94630	**1.81**	0.96485
1.02	0.84614	**1.22**	0.88877	**1.42**	0.92220	**1.62**	0.94738	**1.82**	0.96562
1.03	0.84849	**1.23**	0.89065	**1.43**	0.92364	**1.63**	0.94845	**1.83**	0.96638
1.04	0.85083	**1.24**	0.89251	**1.44**	0.92507	**1.64**	0.94950	**1.84**	0.96712
1.05	0.85314	**1.25**	0.89435	**1.45**	0.92647	**1.65**	0.95053	**1.85**	0.96784
1.06	0.85543	**1.26**	0.89617	**1.46**	0.92785	**1.66**	0.95154	**1.86**	0.96856
1.07	0.85769	**1.27**	0.89796	**1.47**	0.92922	**1.67**	0.95254	**1.87**	0.96926
1.08	0.85993	**1.28**	0.89973	**1.48**	0.93056	**1.68**	0.95352	**1.88**	0.96995
1.09	0.86214	**1.29**	0.90147	**1.49**	0.93189	**1.69**	0.95449	**1.89**	0.97062
1.10	0.86433	**1.30**	0.90320	**1.50**	0.93319	**1.70**	0.95543	**1.90**	0.97128
1.11	0.86650	**1.31**	0.90490	**1.51**	0.93448	**1.71**	0.95637	**1.91**	0.97193
1.12	0.86864	**1.32**	0.90658	**1.52**	0.93574	**1.72**	0.95728	**1.92**	0.97257
1.13	0.87076	**1.33**	0.90824	**1.53**	0.93699	**1.73**	0.95818	**1.93**	0.97320
1.14	0.87286	**1.34**	0.90988	**1.54**	0.93822	**1.74**	0.95907	**1.94**	0.97381
1.15	0.87493	**1.35**	0.91149	**1.55**	0.93943	**1.75**	0.95994	**1.95**	0.97441
1.16	0.87698	**1.36**	0.91309	**1.56**	0.94062	**1.76**	0.96080	**1.96**	0.97500
1.17	0.87900	**1.37**	0.91466	**1.57**	0.94179	**1.77**	0.96164	**1.97**	0.97558
1.18	0.88100	**1.38**	0.91621	**1.58**	0.94295	**1.78**	0.96246	**1.98**	0.97615
1.19	0.88298	**1.39**	0.91774	**1.59**	0.94408	**1.79**	0.96327	**1.99**	0.97670
1.20	0.88493	**1.40**	0.91924	**1.60**	0.94520	**1.80**	0.96407	**2.00**	0.97725

(3)

[ERMFormulaBookAppendixCumulativeNormal3]

The table below tabulates the cumulative distribution function of the unit normal distribution, i.e.

$$N(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}t^2\right) dt$$

for values of x from 2.00 to 3.00:

x	$N(x)$	x	$N(x)$	x	$N(x)$	x	$N(x)$	x	$N(x)$
2.00	0.97725	**2.20**	0.98610	**2.40**	0.99180	**2.60**	0.99534	**2.80**	0.99744
2.01	0.97778	**2.21**	0.98645	**2.41**	0.99202	**2.61**	0.99547	**2.81**	0.99752
2.02	0.97831	**2.22**	0.98679	**2.42**	0.99224	**2.62**	0.99560	**2.82**	0.99760
2.03	0.97882	**2.23**	0.98713	**2.43**	0.99245	**2.63**	0.99573	**2.83**	0.99767
2.04	0.97932	**2.24**	0.98745	**2.44**	0.99266	**2.64**	0.99585	**2.84**	0.99774

2.05	0.97982	**2.25**	0.98778	**2.45**	0.99286	**2.65**	0.99598	**2.85**	0.99781
2.06	0.98030	**2.26**	0.98809	**2.46**	0.99305	**2.66**	0.99609	**2.86**	0.99788
2.07	0.98077	**2.27**	0.98840	**2.47**	0.99324	**2.67**	0.99621	**2.87**	0.99795
2.08	0.98124	**2.28**	0.98870	**2.48**	0.99343	**2.68**	0.99632	**2.88**	0.99801
2.09	0.98169	**2.29**	0.98899	**2.49**	0.99361	**2.69**	0.99643	**2.89**	0.99807
2.10	0.98214	**2.30**	0.98928	**2.50**	0.99379	**2.70**	0.99653	**2.90**	0.99813
2.11	0.98257	**2.31**	0.98956	**2.51**	0.99396	**2.71**	0.99664	**2.91**	0.99819
2.12	0.98300	**2.32**	0.98983	**2.52**	0.99413	**2.72**	0.99674	**2.92**	0.99825
2.13	0.98341	**2.33**	0.99010	**2.53**	0.99430	**2.73**	0.99683	**2.93**	0.99831
2.14	0.98382	**2.34**	0.99036	**2.54**	0.99446	**2.74**	0.99693	**2.94**	0.99836
2.15	0.98422	**2.35**	0.99061	**2.55**	0.99461	**2.75**	0.99702	**2.95**	0.99841
2.16	0.98461	**2.36**	0.99086	**2.56**	0.99477	**2.76**	0.99711	**2.96**	0.99846
2.17	0.98500	**2.37**	0.99111	**2.57**	0.99492	**2.77**	0.99720	**2.97**	0.99851
2.18	0.98537	**2.38**	0.99134	**2.58**	0.99506	**2.78**	0.99728	**2.98**	0.99856
2.19	0.98574	**2.39**	0.99158	**2.59**	0.99520	**2.79**	0.99736	**2.99**	0.99861
2.20	0.98610	**2.40**	0.99180	**2.60**	0.99534	**2.80**	0.99744	**3.00**	0.99865

(4)

[ERMFormulaAppendixCumulativeNormal4]

The table below tabulates the cumulative distribution function of the unit normal distribution, i.e.

$$N(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}t^2\right) dt$$

for values of x from 3.00 to 4.00:

x	$N(x)$	x	$N(x)$	x	$N(x)$	x	$N(x)$	x	$N(x)$
3.00	0.99865	**3.20**	0.99931	**3.40**	0.99966	**3.60**	0.99984	**3.80**	0.99993
3.01	0.99869	**3.21**	0.99934	**3.41**	0.99968	**3.61**	0.99985	**3.81**	0.99993
3.02	0.99874	**3.22**	0.99936	**3.42**	0.99969	**3.62**	0.99985	**3.82**	0.99993
3.03	0.99878	**3.23**	0.99938	**3.43**	0.99970	**3.63**	0.99986	**3.83**	0.99994
3.04	0.99882	**3.24**	0.99940	**3.44**	0.99971	**3.64**	0.99986	**3.84**	0.99994
3.05	0.99886	**3.25**	0.99942	**3.45**	0.99972	**3.65**	0.99987	**3.85**	0.99994
3.06	0.99889	**3.26**	0.99944	**3.46**	0.99973	**3.66**	0.99987	**3.86**	0.99994
3.07	0.99893	**3.27**	0.99946	**3.47**	0.99974	**3.67**	0.99988	**3.87**	0.99995
3.08	0.99896	**3.28**	0.99948	**3.48**	0.99975	**3.68**	0.99988	**3.88**	0.99995
3.09	0.99900	**3.29**	0.99950	**3.49**	0.99976	**3.69**	0.99989	**3.89**	0.99995
3.10	0.99903	**3.30**	0.99952	**3.50**	0.99977	**3.70**	0.99989	**3.90**	0.99995
3.11	0.99906	**3.31**	0.99953	**3.51**	0.99978	**3.71**	0.99990	**3.91**	0.99995
3.12	0.99910	**3.32**	0.99955	**3.52**	0.99978	**3.72**	0.99990	**3.92**	0.99996
3.13	0.99913	**3.33**	0.99957	**3.53**	0.99979	**3.73**	0.99990	**3.93**	0.99996
3.14	0.99916	**3.34**	0.99958	**3.54**	0.99980	**3.74**	0.99991	**3.94**	0.99996
3.15	0.99918	**3.35**	0.99960	**3.55**	0.99981	**3.75**	0.99991	**3.95**	0.99996
3.16	0.99921	**3.36**	0.99961	**3.56**	0.99981	**3.76**	0.99992	**3.96**	0.99996
3.17	0.99924	**3.37**	0.99962	**3.57**	0.99982	**3.77**	0.99992	**3.97**	0.99996
3.18	0.99926	**3.38**	0.99964	**3.58**	0.99983	**3.78**	0.99992	**3.98**	0.99997
3.19	0.99929	**3.39**	0.99965	**3.59**	0.99983	**3.79**	0.99992	**3.99**	0.99997
3.20	0.99931	**3.40**	0.99966	**3.60**	0.99984	**3.80**	0.99993	**4.00**	0.99997

For calculations of specific entries in this table see e.g.:

3.20	0.99931	**3.40**	0.99966	**3.60**	0.99984	**3.80**	0.99993	**4.00**	0.99997

(b) Standard (i.e. unit) normal distribution: Quantile points

[ERMFormulaBookAppendixNormalQuantiles]

The table below gives (percentage) quantile points x for the unit normal distribution defined by the equation:

$$q = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{x} \exp\left(-\frac{1}{2}t^2\right) dt \qquad i.e.\ \ x = N^{-1}(q)$$

q	x	q	x	q	x	q	x	q	x
50%	0.0000	**5.0%**	-1.6449	**2.0%**	-2.0537	**1.0%**	-2.3263	**0.10%**	-3.0902
45%	-0.1257	**4.5%**	-1.6954	**1.9%**	-2.0749	**0.9%**	-2.3656	**0.09%**	-3.1214
40%	-0.2533	**4.0%**	-1.7507	**1.8%**	-2.0969	**0.8%**	-2.4089	**0.08%**	-3.1559
35%	-0.3853	**3.5%**	-1.8119	**1.7%**	-2.1201	**0.7%**	-2.4573	**0.07%**	-3.1947
30%	-0.5244	**3.0%**	-1.8808	**1.6%**	-2.1444	**0.6%**	-2.5121	**0.06%**	-3.2389
25%	-0.6745	**2.5%**	-1.9600	**1.5%**	-2.1701	**0.5%**	-2.5758	**0.05%**	-3.2905
20%	-0.8416	**2.4%**	-1.9774	**1.4%**	-2.1973	**0.4%**	-2.6521	**0.01%**	-3.7190
15%	-1.0364	**2.3%**	-1.9954	**1.3%**	-2.2262	**0.3%**	-2.7478	**0.005%**	-3.8906
10%	-1.2816	**2.2%**	-2.0141	**1.2%**	-2.2571	**0.2%**	-2.8782	**0.001%**	-4.2649
5%	-1.6449	**2.1%**	-2.0335	**1.1%**	-2.2904	**0.1%**	-3.0902	**0.0005%**	-4.4172

For calculations of specific values of this table see e.g.:

5%	-1.6449	**2.1%**	-2.0335	**1.1%**	-2.2904	**0.1%**	-3.0902	**0.0005%**	-4.4172